

WHAT IS CLAIMED IS

1. A method for removing noise from a digital representation of data images and noise images produced by digital scanning of a document, comprising:

(a) performing an object grabbing operation on the digital representation to obtain all objects of the document;

(b) identifying objects that represent essential information of the document and marking them as data objects; and

(c) reconstructing a digital representation of a reduced noise version of the document consisting of all of the marked data objects.

2. A method for removing noise from a digital representation of essential images and noise images produced by digital scanning of a document, comprising:

(a) performing an object grabbing operation on the digital representation to identify adjoining pixels which form objects;

(b) operating a processor so as to identify a data area portion of the digital

6 representation that includes the objects which constitute the essential images;

7 (c) operating the processor to perform a logical ANDing operation between  
8 the data area portion and the digital representation to provide a digital representation of the  
9 essential images without the noise images; and

10 (d) operating the processor to eliminate the noise images located outside of  
11 the data areas from the digital representation to provide another digital representation of the data  
12 images.

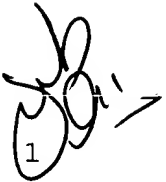
1 3. A method for removing noise from a digital representation of essential images and  
2 noise images produced by digital scanning of a document, comprising:

3 (a) operating a processing system to perform an object grabbing operation on  
4 the digital representation to identify adjoining pixels which form objects;

5 (b) operating the processing system to identify data objects of the digital  
6 representation that constitute essential images;

7 (c) operating the processing system to mark the identified objects as data, and

8 ~~(d) operating the processing system to eliminate all objects not marked as data~~  
9 from the digital representation to provide a reconstructed digital representation of the essential  
10 images without the noise images.

1 

2 4. A method for producing a cleaned-up digital image of a document including  
essential data images and undesired noise images, comprising:

3 (a) digitally scanning the document to produce a first digital representation of  
4 the data images and the noise images;

5 (b) performing a first object grabbing operation on the first digital  
6 representation to identify all object images thereof;

7 (c) determining a skew angle of a straight line having a predetermined  
8 relationship to some objects representative of the essential data images and de-skewing the  
9 document by rotating the first digital representation by an amount equal to the magnitude of the  
10 skew angle to provide a de-skewed first digital representation;

11 (d) performing a second object grabbing operation on the de-skewed first  
12 digital representation to create an object list of all object images of the de-skewed first digital  
13 representation;

14 ~~(e) identifying a portion of the de-skewed first digital representation~~

15 corresponding to a picture region of the document;

16 (f) producing a reduced-resolution representation of the de-skewed first

17 digital representation and performing a second object grabbing operation on the reduced-

18 resolution representation;

19 (g) identifying objects of the reduced-resolution representation representing

20 essential data areas of the document; and

21 (h) constructing the cleaned-up digital image of the document by performing a

22 logical ANDing operation between the picture region and the data areas with the de-skewed first

23 digital representation to eliminate all objects outside of the picture region and the data areas to

24 provide the cleaned-up digital image.

1 5. A method for producing a cleaned-up digital image of a document including


2 essential data images and undesired noise images, comprising:

3 (a) digitally scanning the document to produce a first digital representation of

4 the data images and the noise images;

5 ~~(h) performing a first object grabbing operation on the first digital~~  
6 representation to identify all object images thereof;

7 (c) determining a skew angle of a straight line having a predetermined  
8 relationship to some objects representative of the essential data images and de-skewing the  
9 document by rotating the first digital representation by an amount equal to the magnitude of the  
10 skew angle to provide a de-skewed first digital representation;

11  (d) performing a second object grabbing operation on the de-skewed first  
12 digital representation to create an object list of all object images of the de-skewed first digital  
13 representation;

14 (e) identifying a portion of the de-skewed first digital representation  
15 corresponding to a picture region of the document;

16 (f) identifying objects representing essential data images of the document and  
17 marking the identified objects as data objects; and

18 (g) constructing the cleaned-up digital image of the document by

19 i. combining the objects in the picture region and the marked data  
20 objects, and

21 ii. ~~eliminating all objects not marked as data objects to provide a~~  
22 reconstructed digital representation of the essential images without the noise images.

1 6. The method of Claim 5 including performing the first object grabbing operation  
2 by obtaining serial runlength data from the first digital representation including slices that each  
3 include the length and ending pixel number of a string of connected pixels having a "1" value,  
4 operating line-by-line on the runlength data by means of a decision tree classifier that creates  
5 software objects including a first linked list of a number of further linked lists each of which  
6 contains all of the slices of an object image, entering the slices of the object image into a  
7 software frame in the same order in which the slices are scanned, determining if the object image  
8 can be represented as a trapezoid or as an irregular blob containing all of its slices, fitting the data  
9 in the software frame representing the object image into a decision tree classifier, and operating  
10 the classifier to recognize and assign identifiers to divergences, convergences, and open ends of  
11 the object image and create a new linked list of linked lists representing the object image in the  
12 form of blob records, trapezoid records, divergence records, and convergence records which then  
13 can be conveniently used in subsequent vectorization operations without the need to scan and  
14 recognize data representing the object image.

1 7. The method of Claim 6 including performing the second object grabbing  
2 operation by performing steps which are essentially similar to the steps of the first object  
3 grabbing operation.

1 8. The method of Claim 5 wherein in step (c) the line has the predetermined  
2 relationship to a plurality of text objects in a row of text objects.

1 9. The method of Claim 8 wherein some of the text objects are centered about the  
2 line.

1 10. The method of Claim 8 including building the row of text objects by successively  
2 adding any closest nearby text object to either end of a row initially including a first text object.

1 11. The method of Claim 6 wherein in step (c) the line has the predetermined  
2 relationship to a plurality of geometric objects identified by vectorizing objects larger than a

predetermined size, identifying near-horizontal lines and near-vertical lines of the vectorized objects, and selecting a value of the skew angle which minimizes the mean-square deviation of the near-horizontal and near-vertical lines from orthogonality.

12. The method of Claim 5 including, after step (b), classifying at least a portion of the document as a type including mainly text objects or mainly geometric objects by producing a first reduced-resolution representation of the first digital representation and performing another object grabbing operation on the first reduced-resolution representation to identify objects of the first reduced-resolution representation, determining the numbers of text-character-shaped rectangular objects and geometric objects thereof, respectively, and classifying the document as text type if the number of text-character-shaped objects is greater than the number of geometric objects, and otherwise classifying the document as geometric type.

13. The method of Claim 5 wherein step (f) includes forming a row of text including text objects near to each other and having heights within a predetermined range, and marking all object images produced according to step (b) within the row as data objects.



1        14.    ~~The method of Claim 13 including identifying the object images which are~~  
2        geometric objects and marking them as data objects.

1            15.    The method of claim 14 wherein the identifying of geometric objects includes  
2        identifying only objects which have sufficiently high density and a sufficiently large aspect ratio  
3        as geometric objects.

1            16.    The method of Claim 14 wherein the identifying of geometric objects includes  
2        identifying whole geometric objects by getting a next object image having a size greater than a  
3        predetermined text size, and, if the next object image has a density lower than a predetermined  
4        density, performing a neural network operation to determine if the next object image is a whole  
5        geometry object, and, if the neural network operation determines that the next object image is a  
6        whole geometry object, marking the next object image as a data object.

1            17.    The method of Claim 14 wherein the identifying of geometric objects includes  
2        identifying broken geometry objects by performing a quad tree operation on all object images not  
3        previously identified as either text objects or geometric objects to identify a non-marked object,

4 ~~and repeatedly identifying any nearby non-marked objects of similar shape, to attempt to extend~~  
5 a pattern of similar non-marked objects in opposite directions from the non-marked object, and  
6 marking all objects included in the pattern as data objects.

1 18. The method of Claim 17 including computing a pattern confidence level, and  
2 marking the objects included in the pattern only if the confidence level exceeds a predetermined  
3 level.

1 19. The method of claim 14 including identifying any object images which constitute  
2 dashed lines or dotted lines and marking such identified object images as data objects, by  
3 creating a grid of the wide, short rectangles or a grid of tall, narrow rectangles covering at least a  
4 portion of the document, summing the areas of all dash-sized for dot-sized objects into  
5 appropriate rectangles, eliminating objects in the appropriate rectangles having sufficiently small  
6 area sums, obtaining a histogram all objects in the appropriate rectangles by area and x-  
7 coordinate or y-coordinate, and marking each object having a sufficiently large histogram peak  
8 and located between predetermined coordinate bounds has a dashed object.

1        20.    ~~The method of Claim 5 wherein the producing of the reduced-resolution~~  
2 representation of the de-skewed first digital representation includes representing each tile of four  
3 adjacent pixels of the first digital representation as a single pixel and setting the single pixel to a  
4 "1" state if any of the four adjacent pixels of the tile is at a "1" state and otherwise setting the  
5 single pixel to a "0" state.

21.    The method of Claim 20 including producing a first reduced-resolution  
representation of the first digital representation by representing each tile of four adjacent pixels  
of the reduced-resolution representation as a single pixel and setting that single pixel to a "1"  
state if any of the four adjacent pixels of that tile is at a "1" state and otherwise setting that single  
pixel to a "0" state.

1        22.    A method for producing a cleaned-up digital image of a document including  
2 essential data images and undesired noise images, comprising:

3            (a)    digitally scanning the document to produce a first digital representation of  
4 the data images and the noise images;

5            (b)    operating a processor to perform a first object grabbing operation on the

6 ~~first digital representation to identify all object images thereof;~~

7 (c) operating the processor to determine a skew angle of a straight line having  
8 a predetermined relationship to at least some objects representative of essential data and to de-  
9 skew the document by rotating the first digital representation by an amount equal to the  
10 magnitude of the skew angle to provide a de-skewed first digital representation;

11 (d) operating the processor to perform a second object grabbing operation on  
12 the de-skewed first digital representation to create an object list of all object images of the de-  
13 skewed first digital representation;

14 (e) operating the processor so as to identify a portion of the de-skewed first  
15 digital representation corresponding to a picture region of the document;

16 (f) operating the processor to produce a reduced-resolution representation of  
17 the de-skewed first digital representation and to perform a second object-grabbing operation on  
18 the reduced resolution representation;

19 (g) operating the processor to identify objects of the reduced-resolution  
20 representation representing essential data areas of the document; and

21 (h) ~~constructing the cleaned-up digital image of the document by operating~~  
22 ~~the processor to perform a logical ANDing operation between the picture region and the data~~

23 ~~areas with the de-skewed first digital representation to eliminate all objects outside of the picture~~  
24 ~~region and the data areas to provide the cleaned-up digital image.~~

1 23. A method for producing a cleaned-up digital image of a document including  
2 essential data images and undesired noise images, comprising:

3 (a) digitally scanning the document to produce a first digital representation of  
4 the data images and the noise images;

5 (b) operating a processing system to perform a first object grabbing operation  
6 on the first digital representation to identify all object images thereof;

7 (c) operating the processing system to determine a skew angle of a straight  
8 line having a predetermined relationship to at least some objects representative of essential data  
9 and to de-skew the document by rotating the first digital representation by an amount equal to the  
10 magnitude of the skew angle to provide a de-skewed first digital representation;

11 (d) operating the processing system to perform a second object grabbing  
12 operation on the de-skewed first digital representation to create an object list of all object images  
13 of the de-skewed first digital representation;

14 ~~(e) operating the processing system so as to identify a portion of the de-~~

15 skewed first digital representation corresponding to a picture region of the document;

16 (f) operating the processing system to identify objects representing essential

17 data images of the document and mark the identified objects as data objects; and

18 (g) constructing the cleaned-up digital image of the document by operating

19 the processing system to

20 i. combine the objects in the picture region and the marked data

21 objects to provide the cleaned-up digital image, and

22 ii. eliminate all objects not marked as data objects to provide a

23 reconstructed digital representation of the essential images without the noise images.

1 24. A system for removing noise from a digital representation of data images and

2 noise images produced by digital scanning of a document, comprising:

3 (a) processor means for performing an object grabbing operation on the digital

4 representation to detect all objects of the document;

5 (b) processor means for identifying objects that represent essential  
6 information of the document and marking them as data objects; and

7 (c) processor means for reconstructing a digital representation of a reduced  
8 noise version of the document consisting of all of the marked data objects.

25. A system for removing noise from a digital representation of essential images and  
noise images produced by scanning of a document, comprising:

(a) processor means for performing an object grabbing operation on the digital  
representation to identify adjoining pixels which form objects;

(b) processor means for identifying a data area portion of the digital  
representation that includes the objects which constitute the essential images;

(c) processor means for performing a logical ANDing operation between the  
data area portion and the digital representation to provide a digital representation of the essential  
images without the noise images; and

(d) processor means for eliminating the noise images located outside of the  
data areas from the digital representation to provide another digital representation of data images.

1 ~~26. A system for removing noise from a digital representation of essential images and~~  
2 noise images of a document, comprising:

3 (a) a scanning device for digitally scanning the document to produce the  
4 digital representation of the essential images and the noise images;

5 (b) a processing system including a program stored in the processing system  
6 for performing an object grabbing operation on the digital representation to identify adjoining  
7 pixels which form objects;

8 (c) a program stored in the processing system for initially marking all of the  
9 objects as noise images;

10 (d) a program stored in the processing system for identifying data objects of  
11 the digital representation that constitute essential images and marking the identified data objects  
12 as data objects; and

13 (e) a program stored in the processing system for eliminating all objects not  
14 marked as data objects from the digital representation to provide a reconstructed digital  
15 representation of the essential images without the noise images.



1 27. A system for removing noise from a digital representation of essential images and  
2 noise images of a document obtained by digitally scanning the document, comprising:

3 (a) a processing system including a program stored in the processing system  
4 for performing an object grabbing operation on the digital representation to identify adjoining  
5 pixels which form objects;

6 (b) a program stored in the processing system for identifying data objects of  
7 the digital representation that constitute essential images and marking the identified data objects  
8 as data objects; and

9 (c) a program stored in the processing system for eliminating all objects not  
10 marked as data objects from the digital representation to provide a reconstructed digital  
11 representation of the essential images without the noise images.